



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/732,809	12/11/2003	Johannes Heinecke	324-163	5874
7590	04/30/2007	LOWE HAUPTMAN GILMAN & BERNER, LLP Suite 300 1700 Diagonal Road Alexandria, VA 22314	EXAMINER STOFFREGEN, JOEL	
			ART UNIT 2626	PAPER NUMBER
			MAIL DATE 04/30/2007	DELIVERY MODE PAPER

**Please find below and/or attached an Office communication concerning this application or proceeding.**

The time period for reply, if any, is set in the attached communication.

<b>Office Action Summary</b>	Application No.	Applicant(s)
	10/732,809	HEINECKE, JOHANNES
	Examiner	Art Unit
	Joel Stoffregen	2626

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

1) Responsive to communication(s) filed on 11 December 2003.  
 2a) This action is FINAL.                            2b) This action is non-final.  
 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

4) Claim(s) 1-7 is/are pending in the application.  
 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.  
 5) Claim(s) \_\_\_\_\_ is/are allowed.  
 6) Claim(s) 1-7 is/are rejected.  
 7) Claim(s) \_\_\_\_\_ is/are objected to.  
 8) Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

**Application Papers**

9) The specification is objected to by the Examiner.  
 10) The drawing(s) filed on 11 December 2003 is/are: a) accepted or b) objected to by the Examiner.  
 Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
 Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).  
 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).  
 a) All    b) Some \* c) None of:  
 1. Certified copies of the priority documents have been received.  
 2. Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.  
 3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

1) Notice of References Cited (PTO-892)  
 2) Notice of Draftsperson's Patent Drawing Review (PTO-948)  
 3) Information Disclosure Statement(s) (PTO/SB/08)  
 Paper No(s)/Mail Date 12/11/2003.

4) Interview Summary (PTO-413)  
 Paper No(s)/Mail Date. \_\_\_\_\_.  
 5) Notice of Informal Patent Application  
 6) Other: \_\_\_\_\_.

**DETAILED ACTION**

1. This action is in response to the original application filed on 12/11/2003.
  
2. Claims 1-7 are currently pending in this application. Claim 1 is an independent claim.

***Priority***

3. Receipt is acknowledged of papers submitted under 35 U.S.C. 119(a)-(d), which papers have been placed of record in the file.

***Information Disclosure Statement***

4. The information disclosure statement (IDS) submitted on 12/11/2003 is being considered by the examiner.

***Claim Rejections - 35 USC § 103***

5. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

6. **Claims 1-3 and 6** are rejected under 35 U.S.C. 103(a) as being unpatentable over VAN DEN AKKER (Patent No.: US 6,415,250) in view of WALTON (Patent No.: US 5,392,419).

7. Regarding **claim 1**, VAN DEN AKKER teaches a device for automatically identifying the language of a digital text ("automatic language identification system", column 6, line 40), comprising:

means for prestoring first character strings that occur frequently anywhere respectively in words of a plurality of predetermined languages and characterize said predetermined languages ("probability table 304 includes an entry for every selected word portion 303 that occurs in at least one of the language corpuses 309", column 10, lines 18-20);

means for prestoring second character strings that are atypical anywhere respectively in words of said predetermined languages ("probability table 304 includes... word portions which do not appear in a language corpus 309", column 10, lines 46-49);

means for analyzing words extracted from said digital text thereby constructing for each extracted word all character strings contained in said extracted word ("word portions extracted from the input text 301", column 10, lines 39-40) and having lengths lying between one character and the number of characters in said extracted word ("more or less characters may be included in the predetermined number of characters", column 9, lines 22-23);

means for comparing character strings contained in extracted words to prestored character strings in order to determine scores associated with said predetermined languages ("identification engine 306 searches the probability table 304 for each of the morphologically-significant word portions extracted from the input text 301, summing the relative probability values associated with each language for each of the extracted word portions", column 10, lines 37-42);

means for comparing each of all character strings contained in each said extracted word individually to said first and second prestored character strings of a determined language so that whenever a first character string is found in said extracted word a score associated with said determined language is increased by a first coefficient depending on the position of said first character string found in said extracted word (see column 10, lines 37-42, and FIG. 6, the suffixes are used for scoring, meaning the values are dependent on the position of the characters, since characters from the suffix are used) and whenever a second character string is found in said extracted word said score is decreased by a respective second coefficient that is associated with said found second character string (see FIG. 6, "probability table 304 is altered to include predetermined negative values for those word portions which do not appear in a language corpus 309", column 13, lines 62-64); and

means for comparing said scores for said text associated with said predetermined languages in order to determine the highest of said scores, which identifies the language of said text ("the largest accumulated relative likelihood value,

provided it exceeds zero, identifies the language of the input text 301", column 10, lines 42-44).

However, VAN DEN AKKER does not disclose that the second coefficient increases as the probability of the character string being in the language decreases.

In the same field of language identification, WALTON teaches a second coefficient that increases as the probability of said found second character string in said determined language decreases (see FIG. 6, the skew value of an unknown word increases by multiples of 4, meaning the value increases as words become less likely to be in a particular language, see column 6, lines 20-32).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to use the coefficient modification of WALTON in the language identification system of VAN DEN AKKER in order to better identify the importance of a character string (see WALKER, column 6, lines 5-11).

8. Regarding **claim 2**, VAN DEN AKKER further teaches that a first character string in an extracted word consists of one of the following character strings: a prefix, a pseudo-prefix, a suffix, a pseudo-suffix, an infix, a pseudo-infix ("word portions containing other types of morphemes or portions of morphemes", column 8, lines 66-67, where "affixes [prefixes, suffixes, infixes] are examples of bound morphemes", column 8, lines 9-10).

9. Regarding **claim 3**, VAN DEN AKKER further teaches that said first coefficient of a first character string in said extracted word depends on the frequency of said character string in said determined language ("frequency value indicative of the number of times the selected word portion was found within the corresponding language corpus 309", column 9, lines 36-38).

10. Regarding **claim 6**, VAN DEN AKKER further teaches comparator means for comparing each of said extracted words from said text with frequent words in said determined language and initially listed in storage means so that whenever a frequent word is found in said text said score for said determined language is increased only by a coefficient depending on the frequency of said extracted word in said determined language ("identification engine 306 searches the probability table 304 for each of the morphologically-significant word portions extracted from the input text 301, summing the relative probability values associated with each language for each of the extracted word portions", column 10, lines 37-42).

11. **Claims 4, 5, and 7** are rejected under 35 U.S.C. 103(a) as being unpatentable over VAN DEN AKKER (Patent No.: US 6,415,250) in view of WALTON (Patent No.: US 5,392,419) and in further view of DE CAMPOS (Patent No.: US 6,272,456).

12. Regarding **claim 4**, VAN DEN AKKER and WALTON teach all of the claimed limitations of claim 1.

However, VAN DEN AKKER and WALTON do not disclose that the first coefficient depends on the length of the character string.

In the same field of language identification, DE CAMPOS teaches that said first coefficient of a first character string in said extracted word depends on the length of said character string ("the language ID program module 36 is looking for the longest match to the test letter sequence of letters appearing in the window", column 13, lines 54-56).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to use the length scoring method of DE CAMPOS in the language identification system of VAN DEN AKKER and WALTON in order to "give more weight to the discriminating effect in most larger n-gram profiles" (DE CAMPOS, column 3, lines 36-37).

13. Regarding **claim 5**, VAN DEN AKKER and WALTON teach all of the claimed limitations of claim 1. VAN DEN AKKER further teaches that said first coefficient of a first character string in said extracted word is equal to:

PO (FR) (see FIG. 6, only suffixes are used, meaning PO has a value of zero or one, depending on whether the characters belong to a suffix or not, and the probability value corresponds directly to the frequency FR, see FIG. 4), where PO is a coefficient depending on the position of said first character string in said extracted word (see FIG. 6, suffix 602) and FR is a coefficient depending on the frequency of said first character string in a determined language (see FIG. 4, suffix frequency list generator 406).

However VAN DEN AKKER and WALTON do not disclose a coefficient that also depends on the length of the character string.

In the same field of language identification, DE CAMPOS teaches that said first coefficient of a first character string in said extracted word is equal to:

(FR + LON) ("a score for each language based upon a frequency parameter in the n-gram profiles corresponding to the length of the longest match", column 4, lines 32-34),

where FR is a coefficient depending on the frequency of said first character string in a determined language ("frequency parameter") and LON is a coefficient depending on the length of said first character string ("length of the longest match").

Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to combine the scoring method of DE CAMPOS with the scoring method of VAN DEN AKKER and WALTON in order to "give more weight to the discriminating effect in most larger n-gram profiles" (DE CAMPOS, column 3, lines 36-37).

14. Regarding **claim 7**, VAN DEN AKKER and WALTON teach all of the claimed limitations of claim 1.

However, VAN DEN AKKER and WALTON do not disclose a coefficient depending on the length of a word.

In the same field of language identification, DE CAMPOS teaches comparator means for comparing each of said extracted words from said text with frequent words in

said determined language and initially listed in storage means so that whenever a frequent word is found in said text said score for said determined language is increased only by a coefficient depending on the length of said frequent word ("the language ID program module 36 is looking for the longest match to the test letter sequence of letters appearing in the window", column 13, lines 54-56).

### ***Conclusion***

15. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. A list of the pertinent prior art can be found on the included form PTO-892.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Joel Stoffregen whose telephone number is (571) 270-1454. The examiner can normally be reached on Monday - Friday, 9:00 a.m. - 6:30 p.m..

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Patrick Edouard can be reached on (571) 272-7603. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

JS



PATRICK N. EDOUARD  
SUPERVISORY PATENT EXAMINER